

Abstract

An input speech utterance is segmented into a prefixed time length to make frames, to extract an acoustic feature parameter of each frame. The acoustic feature parameter is frequency-converted by using plural frequency conversion coefficients previously defined. By using all combinations of plural post-conversion feature parameters obtained by the frequency conversion and at least one standard phonemic model, to compute plural similarities or distances of between the post-conversion feature parameters of each of the frames and the standard phonemic model. A frequency converting condition for normalizing the input utterance is decided by using the plural similarities or distances. By using the frequency converting condition, the input utterance is normalized. With this method, even in case there is change of the speaker making a speech utterance, the individual difference of input utterance can be corrected thereby improving the performance of speech recognition.